



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11143483 A**(43) Date of publication of application: **28 . 05 . 99**

(51) Int. Cl. **G10L 3/00**  
**G10L 3/00**  
**G10L 5/04**  
**// G06F 3/16**  
**G06F 3/16**

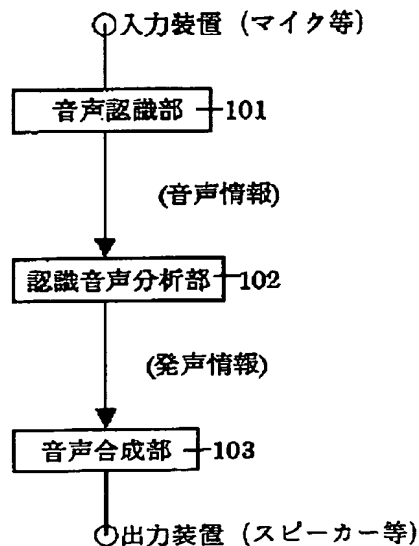
(21) Application number: **10236622**(22) Date of filing: **10 . 08 . 98**(30) Priority: **15 . 08 . 97 JP 09233257**(71) Applicant: **KURITA HIROSHI KATO KEIICHI**(72) Inventor: **KURITA HIROSHI  
KATO KEIICHI**(54) **VOICE GENERATING SYSTEM**

COPYRIGHT: (C)1999,JPO

(57) Abstract:

**PROBLEM TO BE SOLVED:** To allow a user to optionally select various synthetic voices by preparing a voice dictionary generated from the voices of a specific person, connecting and interpolating voice element pieces based on the extracted phoneme code lines, and generating phoneme series.

**SOLUTION:** When the voice of a person is inputted, a speech recognition section 101 recognizes the voice and detects voice information. The detected voice information is analyzed by a recognized voice analysis section 102, and vocalization information is extracted. A voice synthesis section 103 synthesizes the voice signal based on the vocalization information. The voice signal is voice output-converted and outputted. A user can synthesize a voice different from the tone quality of the original person based on the words spoken by the person according to the atmosphere and situation with a voice dictionary independently having features based on the voices of the person and various information. An image matched with the background sound and voice can be selected, and the synthesis of various voices can be enjoyed.



Just Available Copy

Best Available Copy

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-143483

(43) 公開日 平成11年(1999) 5月28日

(51) Int.Cl.<sup>5</sup>

G 1 0 L 3/00

識別記号

5 6 1

5/04

// G 0 6 F 3/16

3 2 0

F I

G 1 0 L 3/00

R

H

5 6 1 D

F

5/04

G 0 6 F 3/16

3 2 0 H

審査請求 未請求 請求項の数35 F D (全 26 頁) 最終頁に続く

(21) 出願番号 特願平10-236622

(22) 出願日 平成10年(1998) 8月10日

(31) 優先権主張番号 特願平9-233257

(32) 優先日 平 9 (1997) 8月15日

(33) 優先権主張国 日本 (J P)

(71) 出願人 393024315

栗田 洋

神奈川県藤沢市片瀬 1 丁目 4 番 13 号 フェリス片瀬 3-201

(71) 出願人 392019433

加藤 圭一

神奈川県川崎市宮前区馬絹1634-1 コスモ宮崎台アバンシード305号

(72) 発明者 栗田 洋

神奈川県藤沢市片瀬 1 丁目 4 番 13 号 フェリス片瀬 3-201

(72) 発明者 加藤 圭一

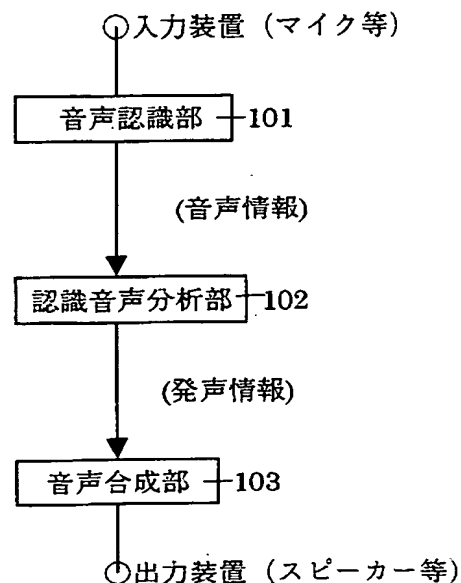
神奈川県川崎市宮前区馬絹1634-1 コスモ宮崎台アバンシード305号

(54) 【発明の名称】 音声発生システム

## (57) 【要約】

【目的】 パソコン、ワープロ、ゲーム機、専用機、N C (ネットワークコンピューター)、STB (セット・トップ・ボックス) などを利用する際の合成音声の発生に係わり、特にユーザーが任意でかつ多様な合成音声を選ぶことが可能な手段を実現するシステム。

【構成】 人の音声を入力しその音声認識を行い、この認識した結果を解析し音韻系列作成のための音韻記号列情報をおよび韻律情報を抽出し、そして特定の人の音声から作成した音声辞書 (音声素片辞書) を準備しておき、前述の抽出した音韻記号列に基づいて音声素片を接続補間し音韻系列を作成する。



## 【特許請求の範囲】

【請求項 1】人の音声を入力し前記音声認識し前記認識した結果を解析し音韻系列作成のための音韻記号列情報を抽出し、特定の人の音声から作成した音声辞書（音声素片辞書）を使用し前記音韻記号列にもとづいて音声素片を接続し音韻系列を作成し前記音韻系列を音声に変換することにより音声合成をすることを特徴とする音声発生システム。

【請求項 2】人の音声を入力し前記音声認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも 1 つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析しテキスト情報を抽出するステップ。

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記テキスト情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音韻系列を音声の出力に変換するステップの各ステップを有することを特徴とする音声発生システム。

【請求項 3】人の音声を入力し前記音声認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも 1 つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報を抽出するステップ

F：前記ユーザーのシステムにおいてはユーザーが任意

の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記音韻系列作成情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音韻系列を音声の出力に変換するステップの各ステップを有することを特徴とする音声発生システム。

【請求項 4】ユーザーに音声合成の元となる音韻系列作成作成のための音韻記号列情報を供給し前記音韻記号列に基づいて音声を発生させるシステムにおいて

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも 1 つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻系列作成するための音韻記号列情報に基づいて前記音声辞書（音声素片辞書）の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音韻系列を音声の出力に変換するステップの各ステップを有することを特徴とする音声発生システム。

【請求項 5】人の音声を入力し前記音声認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

A：特定の人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析しテキスト情報を抽出するステップ

F：前記テキスト情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音韻系列を音声の出力に変換するステップの各ステップを有することを特徴とする音声発生システム。

【請求項 6】人の音声を入力し前記音声認識し前記認

識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

- A：特定の人間の声を基に音声辞書を作成するステップ
  - B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ
  - C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ
  - D：前記入力された音声を認識するステップ
  - E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報を抽出するステップ
  - F：前記音韻系列作成情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ
- ：前記音韻系列を音声の出力に変換するステップ  
の各ステップを有することを特徴とする音声発生システム。

【請求項7】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声を発生させるシステムにおいて

- A：特定の人間の声を基に音声辞書（音声素片辞書）を作成するステップ
  - B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ
  - C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ
  - D：前記音韻記号列情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補完し音韻系列を作成するステップ
  - E：前記音韻系列を音声の出力に変換するステップ
- の各ステップを有することを特徴とする音声発生システム。

【請求項8】人の音声を入力し前記音声を認識し前記認識した結果を解析し音韻系列作成のための音韻記号列情報および韻律情報を抽出し、特定の人の音声から作成した音声辞書（音声素片辞書）を使用し前記音韻記号列にもとづいて音声素片を接続し音韻系列を作成し前記音韻系列に前記韻律を付与することによって音声合成をすることを特徴とする音声発生システム。

【請求項9】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

- A：人間の声を基に音声辞書を作成するステップ
- B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

- D：前記入力された音声を認識するステップ
- E：前記認識した音声を解析しテキスト情報と、韻律情報を抽出するステップ。
- F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

10 G：前記テキスト情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音声素片辞書の音声素片必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

I：前記音声信号を音声の出力に変換するステップ  
の各ステップを有することを特徴とする音声発生システム。

20 【請求項10】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

- A：人間の声を基に音声辞書を作成するステップ
- B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

30 D：前記入力された音声を認識するステップ  
E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報と、韻律情報を抽出するステップ

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記音韻系列作成情報に基づいて前記音声辞書の音声素片辞書の音声素片必要部分を接合し補間し音韻系列を作成するステップ

40 H：前記音声素片辞書の音声素片必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

I：前記音声信号を音声の出力に変換するステップ  
の各ステップを有することを特徴とする音声発生システム。

【請求項11】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声を発生させるシステムにおいて

50 A：人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項12】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声を発生させるシステムにおいて

A：特定の人間の声を基に音声素片および韻律情報からなる音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記音声辞書の前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項13】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号情報及び韻律情報に基づいて音声を発生させるシステムにおいて

A：特定の人間の声を基に音声素片および韻律情報から

なる音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

10 D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列成情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声辞書の韻律情報部をもちいて前記音韻系列の任意の部分の韻律を任意の韻律にするために前記音韻系列のそれぞれの部分に対応する韻律を作成するステップ

20 G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項14】人の音声を入力し前記音声認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させるシステムにおいて

30 A：特定の人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声認識するステップ

E：前記認識した音声を解析しテキスト情報と、韻律情報を抽出するステップ

40 F：前記テキスト情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項15】人の音声を入力し前記音声認識し前記認識した結果を解析し前記解析した結果に基づいて入力

された音声と声質が異なる音声が発生させるシステムにおいて

A：特定の人間の声を基に音声辞書を作成するステップ  
B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声認識するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報と、韻律情報を抽出するステップ

F：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項16】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声が発生させるシステムにおいて

A：特定の人間の声を基に音声辞書を作成するステップ  
B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列系列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項17】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声が発生させるシステムにおいて

A：特定の人間の声を基に音声素片及び韻律情報からなる音声辞書を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ。

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択する

ステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記音声辞書の韻律情報を付加し音声合成のための音声信号を作成するステップ。

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項18】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声が発生させるシステムにおいて

A：特定の人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

E：前記音声辞書の韻律情報部をもちいて前記音韻系列の任意の部分の韻律を任意の韻律にするために前記音韻系列のそれぞれの部分に対応する韻律を作成するステップ。

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【請求項19】前記人間の声を基に音声辞書を作成するプロバイダーと、前記音声辞書を用いて音声信号を作成するユーザーとは分離されていることを特徴とする特許請求範囲第2項、第3項、第4項、第5項、第6項、第7項、第9項、第10項、第11項、第12項、第13項、第14項、第15項、第16項、第17項および第18項記載の音声発生システム。

【請求項20】前記人間の声を基に音声辞書を作成するシステムと、前記音声辞書を用いて音声信号を作成するシステムとは分離されて、それぞれ単独に機能することを特徴とする特許請求範囲第2項、第3項、第4項、第5項、第6項、第7項、第9項、第10項、第11項、第12項、第13項、第14項、第15項、第16項、第17項および第18項記載の音声発生システム。

【請求項21】前記人間の声を基にした音声辞書は音声

素片部分及び韻律部分からなることを特徴とする特許請求範囲第9項記載、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項22】前記人間の声を基にした音声辞書の前記韻律部分は基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含んでいることを特徴とする特許請求範囲第21項記載の音声発生システム。

【請求項23】前記人間の声を基にした音声辞書の各部分は音声素片、基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含んでいることを特徴とする特許請求範囲第9項、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項24】ユーザーは必用に応じてそれぞれの前記人間の声を基にした音声辞書をユーザーの任意の組み合わせで選べることを、前記人間の声を基にした音声辞書から選択した音声辞書の韻律部分の各要素から必用に応じて前記音声の可変可能部分の少なくとも1つをユーザーの任意の選択で可変させることを可能とすることを特徴とする特許請求範囲第9項、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項25】前記可変部分は、声の高さ、発声速度、音量、から選ばれた少なくとも1つ以上であることを特徴とする特許請求範囲第24項記載の音声発生システム。

【請求項26】前記システムにおいて環境音響をデジタル情報に変換し前記環境音響デジタル情報をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じてバックグラウンド音響としての環境音響を前記合成音声に付加することが可能であることを特徴とする特許請求範囲第2項記載、第3項記載、第4項記載、第9項記載、第10項記載、第11項記載、第12項記載及び第13項記載の音声発生システム。

【請求項27】前記システムにおいて環境音響をデジタル情報に変換しユーザーの装置が接続可能な回線に結合されたサーバー上に前記環境音響をデジタル情報に変換した環境音響情報テーブルとして置いておき前記環境音響のデジタル情報をユーザーは必用に応じて前記サーバーに接続し前記バックグラウンド音響としての環境音響

を前記合成音声に付加することが可能であることを特徴とする特許請求範囲第5項記載、第6項記載、第7項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項28】前記人の声を基にした音声辞書の基の人に関連した画像および前記入力に関連した画像から選ばれた少なくとも1つの画像を音声発生時にディスプレイ上に表示することを特徴とする特許請求範囲第2項記載、第3項記載、第4項記載、第5項記載、第6項記載、第7項記載、第9項記載、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項29】前記人の声を基にした音声辞書の基の人に関連した画像および前記入力に関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換しハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記画像情報を付加することが可能であることを特徴とする特許請求範囲第28項記載の音声発生システム。

【請求項30】前記人の声を基にした音声辞書の基の人に関連した画像および前記入力に関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換し画像情報テーブルとしユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記画像情報を付加することが可能であることを特徴とする特許請求範囲第28項記載の音声発生システム。

【請求項31】前記システムにおいて各地方毎に標準化された方言辞書部分を有することを特徴とする特許請求範囲第2項記載、第3項記載、第4項記載、第5項記載、第6項記載、第7項記載、第9項記載、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【請求項32】前記システムにおいて前記方言辞書は解析部、音声合成部に関する音声辞書から選ばれる少なくとも1つに関連づけられることを特徴とする特許請求範囲第23項記載の音声発生システム。

【請求項33】前記方言辞書をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記方言辞書を使用することが可能であることを特徴とする特許請求範囲第31項記載の音声発生システム。

【請求項34】前記方言辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記方言辞書使用

することが可能であることを特徴とする特許請求範囲第31項記載の音声発生システム。

【請求項35】前記システムにおいてユーザーが内容に応じてそれぞれの部分、部分で任意の音声を選べることを特徴とする特許請求範囲第2項記載、第3項記載、第4項記載、第5項記載、第6項記載、第7項記載、第9項記載、第10項記載、第11項記載、第12項記載、第13項記載、第14項記載、第15項記載、第16項記載、第17項記載及び第18項記載の音声発生システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明はパソコン、ワープロ、ゲーム機、専用機、NC（ネットワークコンピュータ）、STB（セット・トップ・ボックス）などを利用する際の合成音声の発生に係わり、特にユーザーが任意でかつ多様な合成音声を選ぶことが可能な手段を実現するシステムに関する。

【0002】

【従来の技術】従来の音声合成に関しては、多種多様の装置及び方法が提案されておりまた構内放送、ゲーム、情報サービス等の様々な分野で応用されている。音声は言語として意味をなす最小単位である音素から構成されている。この音素は波形として表現処理される。その波形の一部あるいは全部が音声の合成に使用される音声素片となる。この音声素片を合成単位として任意の音声を作る合成方式では、各音声素片を結合、補間することによって合成音声が発生させている。

【0003】自然な言語を実現するために各音声素片の接合の工夫も種々の工夫が提案されており人が実際に発声するように自然な発声を実現されている。また文書の意味を解析しその結果を基にして、重要な部分を強調するためにその部分の語調を強めたりする音声合成方法など各種の方法が提案されている。実際の商品としてはドキュメントトーカー、TTSなどの商品などがありテキストファイルを読み上げる音声合成の商品として実際商品化されておりパソコン上で音声合成を実現している。そしてそのほとんどの音声合成はテキストを入力しそのテキストを解析しその解析結果をもとに音声合成を行っている。

【0004】現行カラオケなどで男性の声を女性の声、或いは女性の声を男性の声に変えるといったボイスチェンジャーなどの商品もある。こういった商品は主に周波数変換を行っているため任意の声質の声に変換するといったことはできていなかった。

【0005】

【発明が解決しようとする課題】従来の音声合成では音声合成を行うためにはまずテキストを用意する必要があった。いちいちテキストを用意することなく人の喋る言葉を基に元の人の声質と異なる音声を作成することが

ぞまれていた。また人の言葉を情報化したものから任意の声質の音声を得ることも興味ぶかい。そして人が聞いて違和感のない自然な音声はいつも聞いている人の声をベースにしたものの方が自然なはずである。またいつも同質の声ではメリハリがつかないため、或いは状況などによって聞きたい声質は異なるため、色々な人の声をベースにした音声を状況によって使い分け種々な音声を発声させることも上記と同様に興味深い。例えばある人の声をマイクで拾いその声を他の人の声に変換するといったことが望まれていた。

【0005】人にはそれぞれ声質、基本周波数、アクセント、イントネーション、ポーズの取り方、言葉の一部を伸ばしたりする時間的特徴、固有の方言の特徴、感情により音の高低、音量の変化などの特徴が異なる。こういった人により異なる特徴的要素を加えることが可能となればより臨場感のある音声発声がユーザーの状況に合わせ自由に選べるはずである。

【0006】各地方の方言による音声合成を任意の人の声をベースにしたもの聞きたいなどということも望まれていた。また任意の合成音声を雰囲気或いは状況に応じた任意のバックグラウンド音響の中で聞くということも望まれていた。同様に任意の声質で歌を歌わせることも望まれている。前記ユーザーが任意に選択した声質に関連した画像をディスプレイ上に表示することにより更に臨場感を増すことが可能となる。

【0007】

【課題を解決するための手段】前記目的を達成するために、本発明においては次に示す方法をとる。人の音声を入力しその音声を音声認識を行いこの認識した結果を解析し音韻系列作成のための音韻記号列情報をおよび韻律情報を抽出する。但しこの韻律情報は必要のない場合は抽出しなくてもよい。

【0008】そして特定の人の音声から作成した音声辞書（音声素片辞書）を準備しておき、前述の抽出した音韻記号列に基づいて音声素片を接続補間し音韻系列を作成する。韻律情報を抽出してある場合には音声素片を接続補間した音韻系列にこの韻律情報を付与することによって音声合成をすればよい。

【0009】人の音声を入力しこの音声を認識しこの認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声が発生させる場合においては

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加イン



ストールおよび削除するステップ

D：前記入力された音声を確認するステップ

E：前記認識した音声を解析しテキスト情報を抽出するステップ。

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記テキスト情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0010】人の音声を入力しこの音声を認識し認識した結果を解析し解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記入力された音声を確認するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報を抽出するステップ

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記音韻系列作成情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0011】ユーザーに音声合成の元となる音韻系列を作成するための音韻記号列情報を供給し前記音韻記号列に基づいて音声を発生させる場合においてはにおいて

A：人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻系列作成するための音韻記号列情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0012】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合において

A：特定の人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声を確認するステップ

E：前記認識した音声を解析しテキスト情報を抽出するステップ

F：前記テキスト情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0013】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：特定の人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声を確認するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報を抽出するステップ

F：前記音韻系列作成情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0014】ユーザーに音声合成の元となる音韻系列を作成するための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声を発生させるシステムにおいて

A：特定の人間の声を基に音声辞書（音声素片辞書）を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバ

ー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記音韻記号列情報に基づいて前記音声辞書（音声素片辞書）の音声素片の必要部分を接合し補完し音韻系列を作成するステップ

E：前記音韻系列を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0015】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析しテキスト情報と、韻律情報を抽出するステップ。

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

G：前記テキスト情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

H：前記音声素片辞書の音声素片必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

I：前記音声信号を音声の出力に変換するステップの各ステップを実行することにより達成できる。

【0016】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報と、韻律情報を抽出するステップ

F：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声

辞書を選択するステップ

G：前記音韻系列作成情報に基づいて前記音声辞書の音声素片辞書の音声素片必要部分を接合し補間し音韻系列を作成するステップ

H：前記音声素片辞書の音声素片必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

I：前記音声信号を音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【0017】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声を発生させる場合においては

A：人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

20 C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

30 F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを実行することにより達成できる。

【0018】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声を発生させる場合においては

A：特定の人間の声を基に音声素片および韻律情報からなる音声辞書を作成するステップ

40 B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

50 E：前記音韻記号列情報に基づいて前記音声辞書の音声

素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記音声辞書の前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【0019】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号情報及び韻律情報に基づいて音声が発生させる場合においては

A：特定の人間の声を基に音声素片および韻律情報からなる音声辞書を作成するステップ

B：前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ

C：前記ユーザーに供給された前記人間の声を基にした音声辞書をユーザーのシステムに必要に応じて追加インストールおよび削除するステップ

D：前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声辞書の韻律情報部をもちいて前記音韻系列の任意の部分の韻律を任意の韻律にするために前記音韻系列のそれぞれの部分に対応する韻律を作成するステップ

G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを実行することにより達成できる。

【0020】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：特定の人間の声を基に音声辞書を作成するステップ  
B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析しテキスト情報と、韻律情

報を抽出するステップ

F：前記テキスト情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを実行することにより達成できる。

【0021】人の音声を入力し前記音声を認識し前記認識した結果を解析し前記解析した結果に基づいて入力された音声と声質が異なる音声を発生させる場合においては

A：特定の人間の声を基に音声辞書を作成するステップ  
B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記入力された音声を認識するステップ

E：前記認識した音声を解析し音韻系列作成のための音韻記号列情報と、韻律情報を抽出するステップ

F：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

G：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

H：前記音声信号データを音声の出力に変換するステップ

の各ステップを有することを特徴とする音声発生システム。

【0022】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声が発生させる場合においては

A：特定の人間の声を基に音声辞書を作成するステップ  
B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列系列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステッ

ブ

の各ステップを実行することにより達成できる。

【0023】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報を供給し前記音韻記号列情報に基づいて音声を発生させる場合において

A：特定の人間の声を基に音声素片及び韻律情報からなる音声辞書を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ。

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

E：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記音声辞書の前記韻律情報を付加し音声合成のための音声信号を作成するステップ。

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを実行することにより達成できる。

【0024】ユーザーに音声合成の元となる音韻系列作成のための音韻記号列情報及び韻律情報を供給し前記音韻記号列情報及び韻律情報に基づいて音声を発生させる場合においては

A：特定の人間の声を基に音声辞書を作成するステップ

B：前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C：ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ

D：前記音韻記号列情報に基づいて前記音声辞書の音声素片辞書の音声素片の必要部分を接合し補間し音韻系列を作成するステップ

E：前記音声辞書の韻律情報部をもちいて前記音韻系列の任意の部分の韻律を任意の韻律にするために前記音韻系列のそれぞれの部分に対応する韻律を作成するステップ。

F：前記音声素片辞書の音声素片の必要部分を接合した音韻系列に前記韻律情報を付加し音声合成のための音声信号を作成するステップ

G：前記音声信号データを音声の出力に変換するステップ

の各ステップを実行することにより達成できる。

【0025】そして音声合成する場合には人間の声を基にした音声辞書は必要に応じて音声素片部分、そして基にした人間の声の韻律を使用する場合には韻律部分から構成させればよい。音声素片部分でその人声の質を実現できまた韻律分でその人のしゃべり方の語調等を実現できる。

【0026】基にした人間の声の主な韻律情報を使用する場合には音声辞書の前記韻律部分は必要に応じて基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含んでいる構成とすればよい。

【0027】或いは人間の声を基にした音声辞書の各部分は音声素片、基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分から構成してもよい。

【0028】ユーザーは必用に応じてそれぞれの前述の複数の人間の声を基にした音声辞書をユーザーの任意の組み合わせで選べることで、その人間の声を基にした音声辞書から選択した音声辞書の韻律情報各要素の中から必用に応じてその音声において声の高さ、発声速度、音量を可変可能としこの可変可能部分の3つのうち少なくとも1つをユーザーの任意の選択で可変させればよりユーザーの選択肢が広がり感情等の表現をより強調することができる。

【0029】ユーザーのシステムにおいて各種情報を自分のシステムのハードディスク等にインストール可能な場合にはこのシステムで合成音声のバックグラウンド音響となる環境音響をデジタル情報に変換しこの環境音響デジタル情報をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じてバックグラウンド音響としての環境音響を自分のシステムに導入できるようにして前述の合成音声にこの環境音響を付加してもよい。

【0030】或いはユーザーのシステムが接続可能なサーバー上において各種情報がセット可能な場合には合成音声のバックグラウンド音響となる環境音響をデジタル情報に変換しユーザーの装置が接続可能な回線に結合されたサーバー上にこの環境音響をデジタル情報に変換した環境音響情報テーブルとして置いておきこの環境音響のデジタル情報をユーザーは必用に応じて前記サーバーに接続し前記バックグラウンド音響としての環境音響を前記合成音声に付加してもよい。

【0031】前述の人の声を基にした音声辞書の基の人に関連した画像或いは前述の入力に関連した画像などの画像を音声発生時にディスプレイ上に表示しても効果的である。

【0032】その方法はユーザーのシステムにおいて各種情報を自分のシステムのハードディスク等にインストール可能な場合には前述の人の声を基にした音声辞書の基の人に関連した画像或いは前述の入力に関連した画像などの画像をデジタル情報に変換しハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続

可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前述の画像情報を付加してもよい。

【0033】ユーザーのシステムが接続可能なサーバー上において各種情報がセット可能な場合には前述の人の声を基にした音声辞書の基の人に関連した画像或いは前述の入力に関連した画像などの画像をデジタル情報に変換し画像情報テーブルとしユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前述の画像情報を付加してもよい。

【0034】またこのシステムにおいて各地方毎に標準化された方言辞書部分を設けこの方言辞書はひつよに応じて解析部、音声合成部に関する音声辞書から選ばれる少なくとも1つに関連づけるようにしてもよい。

【0035】ユーザーのシステムにおいて各種情報を自分のシステムのハードディスク等にインストール可能な場合にはこの方言辞書をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じてこの方言辞書を使用できるシステムとすればよい。

【0036】或いはユーザーのシステムが接続可能なサーバー上において各種情報がセット可能な場合にはこの方言辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前述のサーバーに接続しこの方言辞書使用できるようにしてもよい。

【0037】前述のハード媒体はフロッピーディスク、CD-ROM、MO、PD、DVD、メモリーカードなどから最も適当なものを選べばよい。

【0038】前述の回線は一般公衆回線或いは専用回線であり、この回線は有線回線或いは無線回線でこの有線回線は電気信号または光信号であればよい。そして前述の無線回線は電磁波信号及び光信号から選ばれた少なくとも1つの無線回線であるればよい。

【0039】このシステムにおいてユーザーが内容に応じてそれぞれの部分、部分で任意の音声（音声素片辞書、韻律情報等から選択）を選べるようにすれば、すなわち一連の話の流れのそれぞれの部分（合成される音声の部分、部分）で声質、韻律の異なった多用な合成音声を得ることが可能となる。

【0040】そしてユーザーのシステムにおいて各音声辞書、各種情報を自分のシステムのハードディスク等にインストール可能な場合にはこのシステムに使用される装置はパソコン、音発生機能付きワープロ、専用装置などから目的に応じて一番適当なものを選べばよい。

【0041】サーバー上に各音声辞書、各情報を設置している場合にはシステムに使用される装置はパソコン、音発生機能付きワープロ、専用装置、NC（ネットワー

クコンピュータ）、ゲームマシン、STB（セット・トップ・ボックス）から目的に応じて一番適当なものを選べばよい。

【0042】前記システムにおける少なくとも1人以上の声を基にした前記音声辞書、前記環境音響をデジタル情報に変換した情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータを格納したこれらのCD-ROMなどのハード媒体は売り切りとすればよい。

10 【0043】このシステムにおける前述のハード媒体は1人以上の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータを格納しガードをかけ、ユーザーに供給しユーザーの入金を確認後前記ガードを外すシステムとしてもよい。

20 【0044】或いはシステムにおいて前述のハード媒体に格納およびユーザーの接続可能な回線に結合されたサーバー上に格納から選ばれた少なくとも1つの格納データは複数の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータであり、それぞれのデータは独立にガードをかけ、ユーザーの入金を確認後それぞれ独立に前記ガードをはずすようにしてもよい。

30 【0045】前述のシステムにおいてユーザーのシステムが接続可能なサーバー上において各音声辞書、各種情報がセット可能な場合にはユーザーの入金確認により前記サーバー及びサーバー上の辞書、情報の必用部分から選ばれた少なくとも1つのデータにユーザーのIDなどを利用して接続可能とするシステムとしてもよい。

【0047】或いはユーザーのシステムにおいて各音声辞書、各種情報を自分のシステムのハードディスク等にインストール可能な場合にはこのシステムにおいてユーザーの入金により前記サーバーに接続し必要音声辞書部分、情報部分をユーザーの装置にコピー可能としてもよい。そして前述システムにおいてユーザーの装置にコピーされた前記音声辞書、情報が必要に応じて自動消滅するようにしてもよい。

40 【0048】前述のシステムにおいてユーザーのシステムが接続可能なサーバー上において各音声辞書、各種情報がセット可能な場合にはユーザーが前記サーバーに接続し前記サーバー上の音声辞書、情報の使用量に応じて料金を決定し、前記料金をクレジットカード、銀行引き落とし、請求書発行による入金から選ばれた少なくとも1つの決済法をとってもよい。

【0049】

50 【作 用】前述のシステムによれば人の声をベースにした各特徴を独立に有する音声辞書および各情報を使って、各ユーザーが、雰囲気、状況などに合わせて、人の喋る言葉を基に元の人の声質と異なる音声を合成するこ

とが可能となる。各地方の方言、または人により異なる方言的特徴なども加味した任意の音声で実現できる。

【0050】また各音声辞書と同様にバックグラウンド音響もユーザーに容易に供給できるのでユーザーの任意のバックグラウンド音響で任意の音声を聞くことが可能となる。任意の声質の基の人に関連した画像をディスプレイ上に表示することができるのでより臨場感が出る。この合成音声を録音媒体に記録することにより携帯用のオーディオカセットテープにより繰り返して再生させることが可能となり学習などにも役に立つ。

#### 【0051】

【実施例1】図1は実施例1の人の声を認識してその声を基に音声を出力する様子の概略を示したものである。ここで人の声を入力すると音声認識部101で音声を認識し音声情報を検出する。検出された音声情報は認識音声分析部102で分析され、発声情報が抽出される。この発声情報を基に音声合成部103で音声信号が合成される。この音声信号を音声出力変換し出力する。

【0052】図2は音声認識部の一構成例で、従来技術で容易に実現できるものである。入力装置からの音声データは、周波数分析部201で所定の時間（例えば、15から30ms程度）毎に順次周波数分析することにより周波数スペクトルデータを生成してスペクトル特徴抽出部202に送出する。音素検出部203では、特徴抽出部で抽出された特徴をもとにパターンマッチングに基づく音素認識アルゴリズムやスペクトログラムリーディング情報に基づく音素認識アルゴリズムによって音素候補の検出を行う。

【0053】韻律情報抽出部204では、自己相関法などの手法による基本周波数を抽出、さらには母音重心点をサーチが行われる。この母音重心点を基に基本周波数パターンが求められる。母音重心点は（1）有声区間である、（2）振幅の極大点近くである、（3）自己相関係数の安定点で極小点である、などの条件を用いて抽出することができる。このようにして抽出された韻律情報は音声合成部に送られる。

【0054】図3は認識音声分析部の一構成例である。単語候補検出部301では音声認識部から送られてきた音韻記号列候補（認識された音素候補）を基に、あらかじめ用意された単語辞書302を用いて単語照合処理を行う。それぞれ尤度計算されて検出された単語候補は、発話確定部303で発話意図に基づいて作成される重要語辞書304に照らし合わされて再度計算される。その結果として発話意図に沿った発生音声文が生成され、情報抽出部305で音韻記号列が抽出される。

【0055】図4は音声合成部の一構成例である。音韻系列作成部401では、抽出された音韻記号列を基に音声素片辞書405を使用して、それぞれの音声素片を接続補完し音韻系列を作成する。韻律情報部から送られてきた韻律情報は、韻律付加部402において音韻系列に

必要に応じて韻律情報を付加し音声信号作成部403において音声信号を作成する。そしてこの音声信号を音声出力変換部分404において変換し406のスピーカーにおいて音声を出力する。

【0056】次に音声素片辞書の作成について説明する。図5に音声素片辞書の作成のプロチャートを示した。音声提供者に予め決められた文章を朗読してもらう。この際音声提供者にはヘッドホンを装着してもらい、このヘッドフォンを通じて文書の朗読のタイミングを指示する。音声提供者はこの指示に基づいて前述の予め決められた文章を朗読していく。この朗読した音声を図5の501に示すようにアナログ録音する。

【0057】次に502においてアナログの音声をデジタル変換していく。このデジタル変換された音声データを503に示すように時間分割処理を行う。次に504に示すように予め決められた文書の音素と音声提供者の音素の対応付けを行う。この対応付けを行った結果を元に各音素にラベリングを行う。この結果を音素ラベリングテーブル508に格納する。音素ラベリングテーブルおよび予め作成してある音声素片切り出しテーブル509に従い506で音声素片を切り出す。切り出された音声素片は507で示された格納処理部において辞書化処理を行い音声素片辞書510に格納する。

【0058】この音声素片辞書を図4の音声素片辞書405にマスクROMの形で格納しておく。そうすることによって上述したように人の音声認識しその音声を基に入力した音声とは声質の異なる声を出力することが可能となる。本実施例では韻律情報を付加したがしなくてもよい。その場合音韻記号列をそのまま音声に変換すればよい。

#### 【0059】

【実施例2】次に他の実施例で説明する。図6に示すように、実施例1と同じ方法で人の声から音声素片辞書を作成し、この音声素片辞書を601のCD-ROMに格納するそしてこのCD-ROMをユーザーに供給しユーザーのシステム602にハードディスク603この音声素片辞書をインストールすればよい。ここでユーザーは任意の人の音声素片辞書（A氏、B氏そしてC氏）を自分のシステムにインストールしておく。

【0060】音声素片辞書を一般電話回線を通じてユーザー供給する場合には図7に示すように音声素片辞書を701のサーバーシステムのハードディスク702にセットしユーザーは一般電話回線を通じて音声素片辞書を取り込み自分のシステム703のハードディスク704にインストールすればよい。

【0061】図8に本実施例の人の音声認識しその音声を基に音声出力をする様子の概略プロチャートを示す。まず人の声を入力し音声認識部801において音声を認識し、次に802の認識音声分析部で認識した音声から必要な情報を抽出する。ここでは発生音声文を生成し、

さらにこの生成した発声音声文から音韻記号列情報と韻律情報を抽出する。そして、発話意図から感情成分を抽出する。次に音声辞書選択部803においてユーザーの任意の人から作成した音声辞書を音声辞書（音声素片辞書）804から選択する。ここではA氏の声から作成した音声素片辞書Aを選んだ。

【0062】次に809の選択したA氏の声から作成した音声素片辞書を使用して音韻記号列から音韻系列を805の音韻系列作成部で作成する。次に806の韻律付加部でこの音韻記号列に韻律を付加する。さらに、この韻律を付加した音韻系列に感情成分による補正（例えば、発話意図に怒りがある場合、発話速度を速くするなど）を加えた上で、807の音声信号作成部において音声信号に変換し808の音声出力変換部で音声出力にし809のスピーカーにて音声を出力する。

【0063】この場合の認識音声分析部の働きを図3を使って詳しく説明する。実施例1の場合と同様に認識された音素候補を基に単語照合処理を行う。さらに、発話意図確定部303において重要単語辞書304と照らし合わせて再度計算することにより、発話意図に沿った発声音声文を生成することができる。実施例1では、この発声音声文の読み上げ情報である音韻記号列のみを利用するが、ここでは、この発声音声文をテキスト情報として情報抽出部305に送出してテキスト解析し、フレーズ成分などの韻律情報、さらには感情成分を抽出する。

【0064】人の音声を基にした音声辞書に音声素片だけではなく韻律部分を作成しておいてもよい。その内容は図9に示す。ここでそれぞれ901は音声素片辞書であり902は韻律辞書である。さらにこの韻律辞書部分は903の基本周波数部分、904の発音される個々の語が持つ強さまたは高さの配置であるアクセント部分、905の話をするとき話し手の上がり下がり調子であるイントネーション部分、906の時間的特徴部分、907の声の大きさを表す音量の特徴、908の方言的特徴、909のポーズの取り方の特徴、910の感情による発声の特徴の独立部分を備えている。

【0065】韻律辞書の作成方法について説明する。韻律辞書も音声素片辞書と同様に予め決められた文章を音声提供者に読んでもらう。但し音声素片辞書作成の時と異なるのは読む際のタイミングの指示を音声提供者に行わない。また興奮した場合、悲しい場合、怒った場合等の感情の起伏を予め決められた文書の必用部分に添付しておく。そして韻律辞書の場合には標準の韻律を定めておき各音声提供者の辞書は標準との差の形で作成する。この時予め標準として定めた男性の声と女性の声の韻律辞書（標準の韻律辞書）を作成しておく。但し標準の韻律辞書の感情部分は抑揚のない状態での基本周波数、音量、発話速度を格納しておく。

【0066】図10の(a)に標準の音声波のアウトライン、(b)に音声提供者Aの音声波のアウトラインを

示す。基本周波数の採取については音声提供者に抑揚のない平板な状態で読んで貰う。基本周波数については標準の基本周波数に対して(b)の音声提供者Aの音声波をそれぞれ予め設定された語についての基本周波数を比較し、その結果を音声辞書の韻律部の基本周波数の部分に格納する。

【0067】次にアクセントについては音声提供者にいつも話している状態で読んでもらう。予めそれぞれ設定された各単語の各部分のピーク部分の位置、高さおよび基本周波数を標準の音声波および音声提供者Aの音声波を比較し標準し、この結果の差を音声辞書の韻律部のアクセントの部分に格納する。ここでピークの音声の位置は図10においては標準のものでは1002であり、音声提供者Aにおいては1012である。また高さ（音声の強さ大きさ）は標準のものでは1001であり、音声提供者Aにおいては1011である。

【0067】イントネーションについては予め決められた文を音声提供者にいつも話している状態で読んでもらう。そして基本周波数の変化を標準のものと比較しその結果の差を音声辞書の韻律部のイントネーションの部分に格納する。図11にその様子を示した。ここで図10の(a)の標準の基本周波数の変化を図11(a)に示すようにdFとして縦軸にとった。同様に図10の

(b)の音声提供者Aの基本周波数の変化を図11(b)に示すようにdFとして縦軸にとった。

【0068】時間的特徴については予め決められた文章中の各音および代表的単語の継続時間時間、てにおはおよび接続詞の最後の音の継続時間そして文章全体を読む早さを比較する。前記と同様に標準のものと比較しその結果の差を音声辞書の韻律部の時間的特徴の部分に格納する。韻律部分（辞書）も音声素片と同様の方法でユーザーのシステムにインストールすればよい。

【0069】前述のように作成した音声辞書の韻律部分を使い前述の音韻系列に対応した韻律情報を作成しこれを音韻系列に付加してもよい。本実施例で説明すればA氏の声から作成した音声素片辞書を使用して音韻記号列から音韻系列を作成するとともにA氏の声から作成した韻律辞書を使用し上述の音韻系列に対応する韻律情報を作成しこれを806の韻律付加部において付加すればよい。

【0070】方言を盛り込む場合について述べる。基本的な流れは図8と同様であるが、802の認識音声分析部から806の韻律付加部までについて方言を盛り込む場合について更に詳しく説明する。この部分は図13に示したが、ここで1308は地方毎に標準化して作成した方言辞書であり単語の置き換えテーブル及び各言葉の韻律の辞書である。図12に方言辞書の構造をしめした。

【0071】1301の認識音声分析部においてテキストを解析し1302の方言辞書選択部で方言辞書を選択

する。そして必要に応じて 1303 の単語置き換え部で単語を置き換え音韻記号列を生成し 1304 の音声辞書選択部で音声辞書の選択を行う。そして 1305 の音韻系列作成部で 1310 の音声辞書の音声素片部分を使い A 氏の音韻系列を作成する。次に 1308 方言辞書及び音声辞書の韻律部分 1309 を使い音韻系列に対応した A 氏の韻律情報を 1307 の韻律作成部において作成する。この韻律情報を韻律付加部 1306 で付加すればよい。

【0072】本実施例ではユーザーのシステムがパーソナルコンピューターのようにそのシステムの中でハードディスクの容量が十分で複数の音声辞書等が格納可能な場合であるが、例えばネットワークコンピュータなどのように音声辞書が格納できない場合には一般電話回線を使いサーバー上の音声辞書を必要に応じてその都度使用し音声を作成してあげればよい。

### 【0073】

【実施例 3】図 14 に本実施例の人の音声認識しその音声に基づき音声出力をする様子の概略フロチャートを示す。ここで実施例 2 と同様の方法で音声辞書を作成し、CD-ROM などの媒体に格納しユーザーに配布するか或いは公衆回線を通じてユーザーに配布しユーザーは任意の人の音声素片辞書（A 氏、B 氏そして C 氏）を自分のシステムにインストールする。

【0074】まず人の声を入力し 1401 の音声認識部において音声を認識し、次に 1402 の認識音声分析部で認識した音声から情報を抽出する。ここでは音韻系列作成のための音韻記号列情報および韻律情報を抽出する。次に 1403 の音声辞書（音声素片辞書）選択部においてユーザーの任意の人から作成した音声辞書を 1408 の音声辞書（音声素片辞書）から選択する。ここでは B 氏の声から作成した音声素片辞書 B を選んだ。

【0075】次に 1408 の選択した B 氏の声から作成した音声素片辞書を使用して音韻記号列情報より 1404 の音韻系列作成部において各音声素片を接続補完し音韻系列を作成する。次に 1405 の韻律付加部でこの音韻系列に韻律を付加する。そしてこの韻律を付加した音韻系列を 1406 の音声信号作成部において音声信号に変換し、1407 の音声出力変換部で音声出力に変換し 1409 のスピーカーにて音声を出力する。

【0076】実施例 2 と同様前述のように作成した音声辞書の韻律部分を使い前述の音韻系列に対応した韻律情報を作成しこれを音韻系列に付加してもよい。本実施例で説明すれば B 氏の声から作成した音声素片辞書を使用して音韻記号列から音韻系列を作成するとともに B 氏の声から作成した韻律辞書を使用し上述の音韻系列に対応する韻律情報を作成しこれを 1405 の韻律付加部において付加すればよい。実施例 2 と同様に認識音声分析部および音声認識部において抽出した韻律情報を韻律付加部において韻律情報付加しているが、音韻系列を作成し

そのまま音声信号に変換しても、また標準の韻律情報を付加するか、或いは音声辞書の韻律情報すなわち B 氏の話す様子から作成した韻律情報を付加してもよい

【0077】方言を盛り込む場合について述べる。基本的な流れは図 14 と同様であるが、1402 の認識音声分析部から 1405 の韻律付加の部分までについて方言を盛り込む場合について更に詳しく説明する。これは図 15 に示したが、ここで 1508 は地方毎に標準化して作成した方言辞書であり各単語の音韻記号列の置き換えテーブル及び各単語の韻律の辞書であり図 12 に示したとおりである。

【0078】1501 の認識音声分析部において情報を抽出し 1502 の方言辞書選択部で方言辞書を選択するそして必要に応じて 1503 の単語の音韻記号列置き換え部で音韻記号列を置き換え新しい音韻記号列を生成し 1504 の音声辞書選択部で音声辞書の選択を行う。そして 1505 の音韻系列作成部で 1510 の音声辞書の音声素片部分を使い B 氏の音韻系列を作成する。次に 1508 及び音声辞書の韻律部分 1509 を使い音韻記号列に対応した B 氏の韻律情報を 1507 の韻律作成部において作成する。この韻律情報を韻律付加部 1506 で付加すればよい。

【0079】また図 16 に示すように音声のある部分のみを他の韻律に変えてもよい。この図は時間とともに語調（音の高さ話すスピードそしてポーズ等を意味する）及び声質の変化させる様子を示している。ここでは 1601 部では入力音声の韻律（語調）、1602 部では B 氏の韻律（語調）B 氏の声質、1603 では C 氏の韻律（語調）B 氏の声質としている。これを実現するためには基本的流れは図 14 と同様であるが認識音声分析部 1402 から韻律付加部 1405 の部分についてさらに詳しく述べる。これを図 17 に示すが、1701 は認識音声分析部、1702 は音声辞書及びその音声素片部分、韻律部分の選択部である。そして 1705 は韻律情報作成部である。

【0080】まず図 16 の 1601 は図 17 の音韻系列作成部 1703 において 1707 の音声辞書の音声素片部分 1707 の A 氏の音声素片を使用し韻律は認識音声分析部 1701 で抽出した韻律情報をそのまま使用し韻律付加部 1705 において付加する。次に図 16 の 1602 部は図 17 の音韻系列作成部 1703 において 1707 の音声辞書の音声素片部分 1707 の B 氏の音声素片を使用し、韻律は韻律作成部 1706 において 1708 の音声辞書の韻律部分の B 氏の部分を使用して音韻系列に対応させ作成した韻律情報を韻律付加部 1705 において付加する。次に図 16 の 1603 部は、図 17 の音韻系列作成部 1703 において音声辞書 1707 の音声素片部分から B 氏の音声素片を使用し、韻律は韻律作成部 1706 において 1708 の音声辞書の韻律部分の B 氏の部分をしようして音韻系列に対応させ作成した韻



律情報を韻律付加部 1705 において付加すればよい。

【0081】実施例 2 と同様本実施例ではユーザーのシステムがパーソナルコンピュータのようにそのシステムの中でハードディスクの容量が十分で複数の音声辞書等が格納可能な場合であるが、例えばネットワークコンピュータなどのように音声辞書が格納できない場合には一般電話回線を使いサーバー上の音声辞書を必要に応じてその都度使用し音声合成してあげばよい。

【0082】

【実施例 4】図 18 に本実施例の音韻系列を作成するための音韻記号列を韻律付きでユーザーに配布し音声出力をする様子の概略フロチャートを示す。ここで前の実施例と同様に音声素片辞書は前述と同じ方法で作成し、CD-ROM などの媒体に格納しユーザーに配布するか或いは公衆回線を通じてユーザーに配布しユーザーは任意の人の音声素片辞書（A 氏、B 氏そして C 氏）を自分のシステムにインストールする。

【0083】またユーザーに音韻系列を作成するための音韻記号列を韻律付きで配布するのは通信回線を通じて送信する。この送信された韻律付き音韻記号列を 1801 の情報分割部において音韻記号列及び韻律情報に分割する。次に音声辞書の選択部 1802 においてユーザーの任意の人から作成した音声辞書を 1808 の音声辞書部（音声素片辞書）から選択する。ここでは B 氏の声から作成した音声素片辞書 B を選んだ。

【0084】次に 1808 の選択した B 氏の声から作成した音声素片辞書を使用して前述の音韻記号列情報より 1803 の音韻系列作成部において各音声素片を接続補間し音韻系列を作成する。次に 1804 の韻律付加部でこの音韻系列に前述の情報分割部 1801 において分割した韻律情報を付加する。そしてこの韻律を付加した音韻系列を 1805 の音声信号作成部において音声信号に変換し 1806 の音声出力変換部で音声出力にし 1809 のスピーカーによって音声を出力する。

【0085】本実施例ではユーザーに音韻系列を作成するための音韻記号列情報を韻律情報付きで配布するのは通信回線を通じて送信しているが例えば無線放送に乗せてもよい。また書籍の文章を音韻系列を作成するための音韻記号列情報を韻律付き情報に加工して CD-ROM などのハード媒体に格納してもよい。音韻記号列情報と韻律情報を別々に配布してもよい。

【0086】情報分割部において抽出した韻律情報を韻律付加部において韻律情報付加しているが、音韻系列を作成しそのまま音声信号に変換しても、また標準の韻律情報を付加するか、或いは音声辞書の韻律情報すなわち B 氏の話す様子から作成した韻律情報を付加してもよい。

【0087】本実施例においてはユーザーに音韻系列を作成するための音韻記号列を韻律付きで配布するのは通信回線を通じて送信したがこれに限られるものではなく

ハード媒体に格納して配布してもよい。もちろん実施例 2, 3, 4 に述べた組み合わせに限られるものではない。他の組み合わせも本発明の範囲内で用途目的に応じて自由である。次に前述の実施例に付加する情報について述べる。

【0088】任意の人の声で音声合成したものに任意のバックグラウンド音響を付加するとさらに臨場感が増す。このために環境音響についても環境音響情報テーブルを作成する。前述したように海の波の音、駅構内の音、高速道路の脇の車の風きり音、効果音楽などを図 19 に示したようにアナログ録音したものを合成音声に付加できる形にデジタル変換し環境音響情報テーブル 1901 に格納すればよい。

【0089】前述の環境音響情報テーブルを前述の音声辞書と同様にハード媒体或いは公衆回線を通じてユーザーのシステムに供給しユーザーのシステムにインストールする。そして図 20 に示すように音声信号に環境音響付加部 2001 において 2003 の環境音響情報テーブルから環境音響を付加し音変換部 2002 においてこれを音に変換し 2004 のスピーカーで出力する。

【0090】本実施例ではユーザーのシステムがパーソナルコンピュータのようにそのシステムの中でハードディスクの容量が十分で環境音響情報テーブルが格納可能な場合であるが、例えばネットワークコンピュータなどのように環境音響情報テーブルが格納できない場合には一般電話回線等を使いその都度サーバー上の環境音響情報テーブルにアクセスしこれを使用すればよい。

【0091】任意の人の音声に関連した画像あるいは入力に関連した画像、又はその双方を音声発生時にディスプレイ上に表示することも効果的である。これを実現するためには、画像情報テーブルを必用とする。これも前述した環境情報テーブルと同様に作成すればよい。例えば任意の人の声の基の音声提供者がアニメーションの主人公の声優であれば、そのアニメーションの主人公の話をしている状態の画像を取り込み画像情報テーブルに格納すればよい。

【0092】また音声提供者自身をスチルカメラでとり静止画像として取り込んでもよいし、またビデオカメラで話をしている所を取り込んでもよく、これを図 21 に示すように音声発生データに付加できる形に変換し画像情報テーブルに格納すればよい。入力に関連した画像も同様に、その入力の想定が列車の旅の場面であれば列車をスチルカメラあるいはビデオカメラで撮影しておき音声信号データに付加できる形に変換し画像情報テーブルに格納すればよい。

【0093】前述の画像情報テーブルを前述の音声辞書と同様にハード媒体或いは公衆回線を通じてユーザーのシステムに供給しユーザーのシステムにインストールする。そして図 22 に示すように音声信号に画像付加部 2201 において 2204 の画像情報テーブルから画像情

報を付加し音変換部2202においてこれを音に変換し2205のスピーカで出力し、2203の画像処理部で画像処理を行い2206のディスプレイにおいて表示する。

【0094】次に各辞書及び各情報のユーザーへの供給方法および課金の方法についてのシステムを説明する。ユーザーのシステムがパソコンの場合を説明したが音声発生ソフトおよび各辞書が回線に結合されたサーバー上に常駐させ、必用に応じてユーザーの装置から前記サーバーに接続して前記音声発生ソフト、各辞書、各情報テーブルを使用するNC（ネットワークコンピュータ）的使用の場合と異なるため分けて説明する。まずユーザーのシステムがパソコンの場合には、大きく分けて2つの方法があり、1つはハード媒体で供給する方法であり、もう1つは回線で供給する方法である。

【0095】前者においてはまず音声辞書、環境音響情報テーブル、画像情報テーブルの各辞書、各情報テーブルを一切ガードをかけずにCD-ROMなどの媒体に格納する。そしてこのCD-ROM媒体をユーザーに売り切りで販売する方法がある。またこの音声発生ソフトおよび各辞書それぞれに異なった暗証番号でガードが外れるようにしてインストールする。そしてこの場合CD-ROM媒体を低価格で販売するか、あるいは無料で配布しユーザーは必用部分のみの代金を払い、供給者はこの代金の入金を確認したらユーザーに必用部分の暗証番号を発行する。

【0096】この暗証番号発行方法は公衆回線を通じてパソコン通信で配布すればよい。もちろん郵送、電話で告げるなどの方法でもよい。ユーザーはこの暗証番号を利用してCD-ROM媒体の必用部分のガードをはずしユーザーのパソコンシステムのハードディスクにインストールする。

【0097】公衆電話回線を通じて供給してもよい。各辞書、買う情報テーブルを公衆電話回線に結合したサーバー上に常駐させておく。ユーザーは必用とする辞書、情報テーブルを前述の回線を通じて入手すればよい。この場合代金の回収は前述のCD-ROMの場合と同様に音声発生ソフトおよび各辞書それぞれに異なった暗証番号でガードをかけておきサーバー上に常駐させておけばよい。そしてユーザーは自由に前記音声発生ソフト、各辞書、各情報テーブルの必用部分を前記回線を通じて自分の装置に取り込む。前述のCD-ROMの場合と同様な方法でガードをはずしユーザーは自分のシステムに必用部分をインストールすればよい。

【0098】ユーザーのシステムに無償あるいは有償でCD-ROM媒体あるいは公衆電話回線を通じてインストールされた各辞書、情報テーブルの必用部分は試用期間を設けておき試用期間がすぎたら自動消滅するように設定しておいてもよい。

【0099】また他の方法として前記回線に結合された

サーバー上の各辞書、各情報テーブルにそれぞれ独立に接続のためのゲートを設ける。このゲートを通過するためにはそれぞれ暗証番号を入力することが必用である構造とする。ユーザーは必用部分のみの代金を払い、供給者はこの代金の入金を確認したらユーザーに必用部分の暗証番号を発行する。ユーザーは前記回線を通じてサーバーに接続し各辞書、各情報テーブルの必用部分に前記暗証番号を使い接続する。そして自分のシステムに必用部分をインストールすればよい。

【0100】次にNC（ネットワークコンピュータ）的使用の場合について説明する。この場合音声発生ソフトおよび各辞書は公衆電話回線に結合されたサーバー上のみ存在しているわけであるから、音を発生させるためにはユーザーの装置から常に前記回線に結合されたサーバーに接続するひつようがある。この場合にはサーバーに接続するためのゲートを設けておき、供給者はユーザーの入金を確認したらそのユーザーが前記サーバーに接続可能な状態にするために前記ゲートを通過できるようにすればよい。また前述のように暗証番号を発行しその暗証番号によりゲートを通過できる構造でもよい。

【0101】このゲートを通過可能な状態はユーザーの入金状態により通過不能としてもよい。例えばこのゲートの通過可能状態の期限を設定し、その期限が過ぎた場合通過不能とし、新たに入金があれば再びゲートを通過可能とすればよい。

【0102】他の方法でもよく供給者はユーザーの使用量に応じて請求してもよい。例えば各辞書、各情報テーブルへのアクセスタイムを記録しておき、このアクセスタイムに応じてユーザーの銀行口座から必用金額を引き落とせばよい。

#### 【0103】

【発明の効果】以上説明したように本発明によれば、人間の声を基にした音声辞書を容易にユーザーに供給することが可能となり、ユーザーのシステムにおいては、ユーザーの任意の音声およびバックグラウンド音響、音声に合致した画像等が選択できるため多種多様の音声合成を楽しむことが可能となる。

#### 【図面の簡単な説明】

【図1】実施例1のフロチャート

【図2】音声認識部の構成

【図3】認識音声分析部の構成

【図4】音声合成部の構成

【図5】音声素片辞書の作成のフローチャート

【図6】音声辞書の供給のフローチャート

【図7】公衆回線による音声辞書供給の図

【図8】実施例2の全体を示すフローチャート

【図9】音声辞書の構成

【図10】（a）標準の音声波のアウトライン

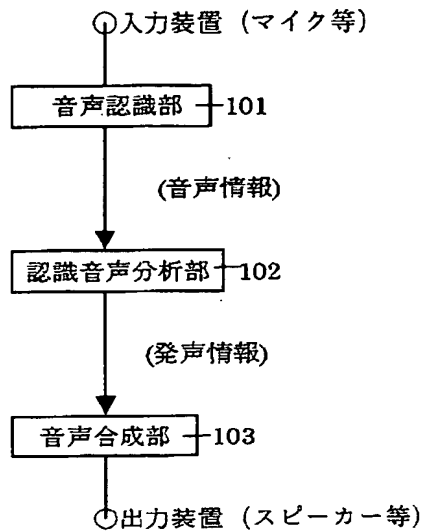
（b）音声提供者Aの音声波のアウトライン

【図11】（a）標準の音声波の基本周波数変化

33

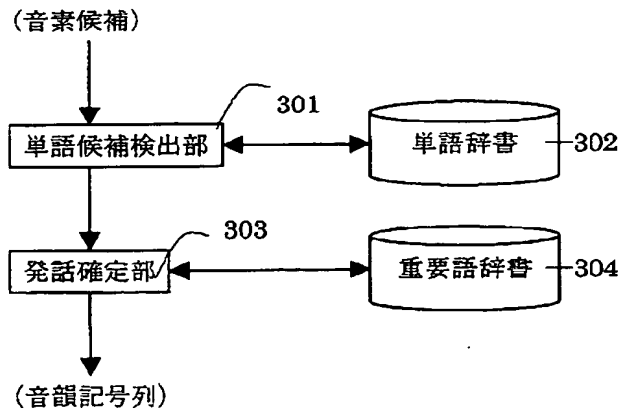
- (b) 音声提供者Aの音声波の基本周波数の変化  
 【図12】 方言辞書の構成  
 【図13】 方言を盛り込む場合のフローチャート  
 【図14】 実施例3の全体のフローチャート  
 【図15】 実施例3に方言を盛り込む場合  
 【図16】 時間と出力したい音声の種類

【図1】



【図3】

認識音声分析部の構成

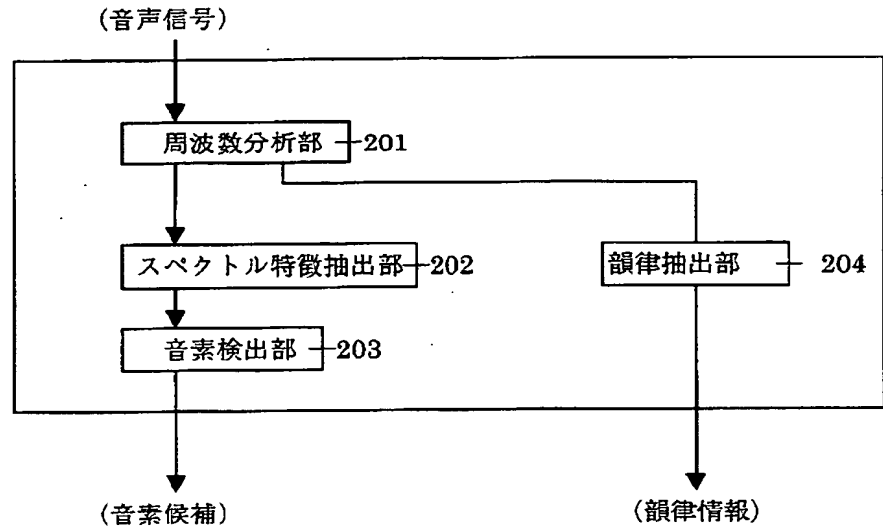


34

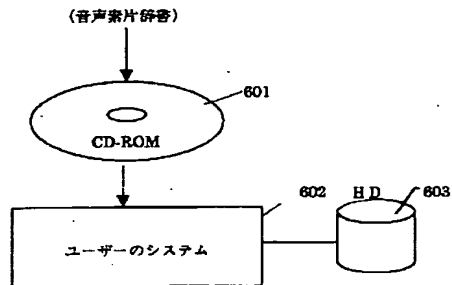
- 【図17】 実施例3に複数の音声を盛り込む場合  
 【図18】 実施例4のフローチャート  
 【図19】 環境情報テーブルの作成  
 【図20】 環境音響情報の付加  
 【図21】 画像情報テーブルの作成  
 【図22】 画像情報の付加

【図2】

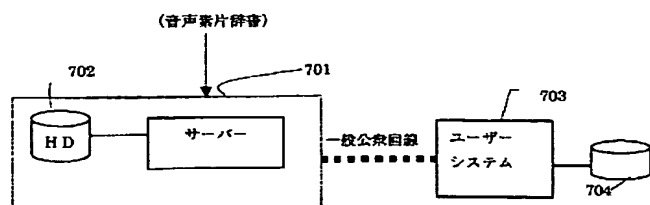
音声認識部の構成



【図6】

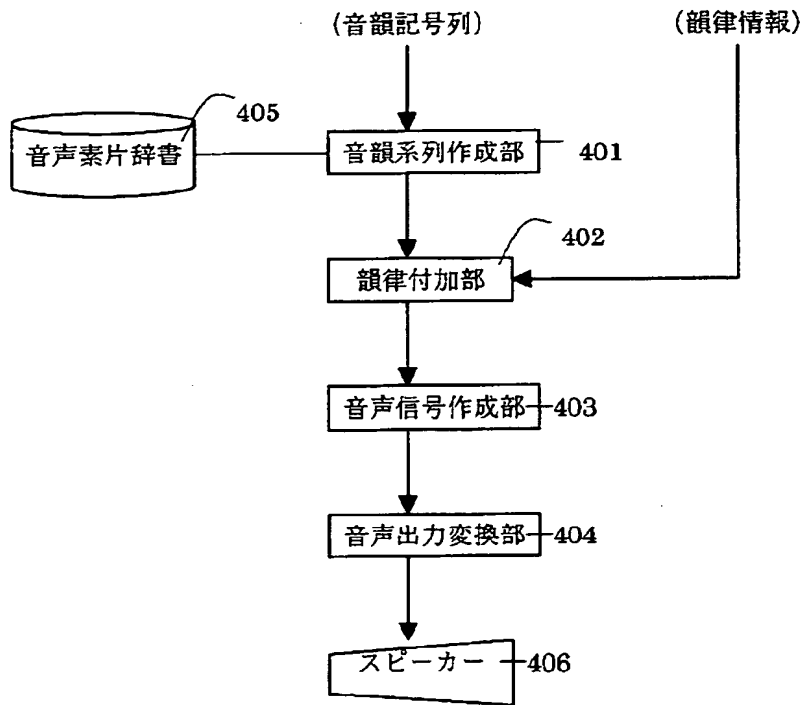


【図7】



【図4】

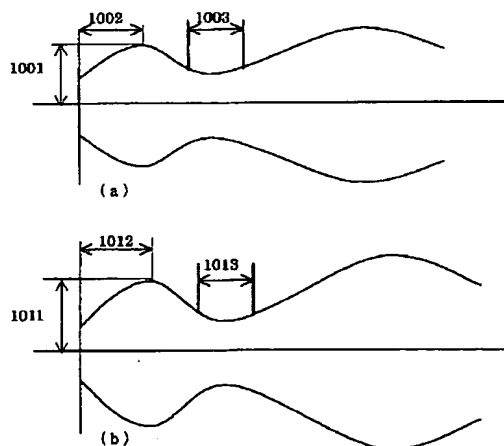
## 音声合成部の構成



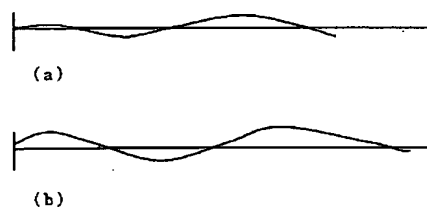
【図9】

音声 素片 辞書	韻律 辞書	基本周波数	903
		アクセント	904
		イントネーション	905
		時間的特徴	906
		音量の特徴	907
		方言的特徴	908
		ポーズの取り方の特徴	909
		感情による発声の特徴	910

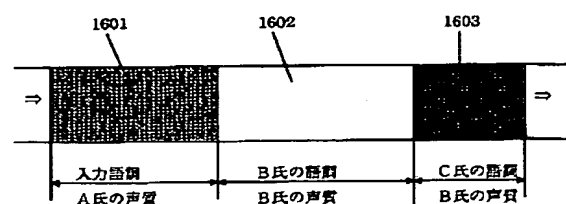
【図10】



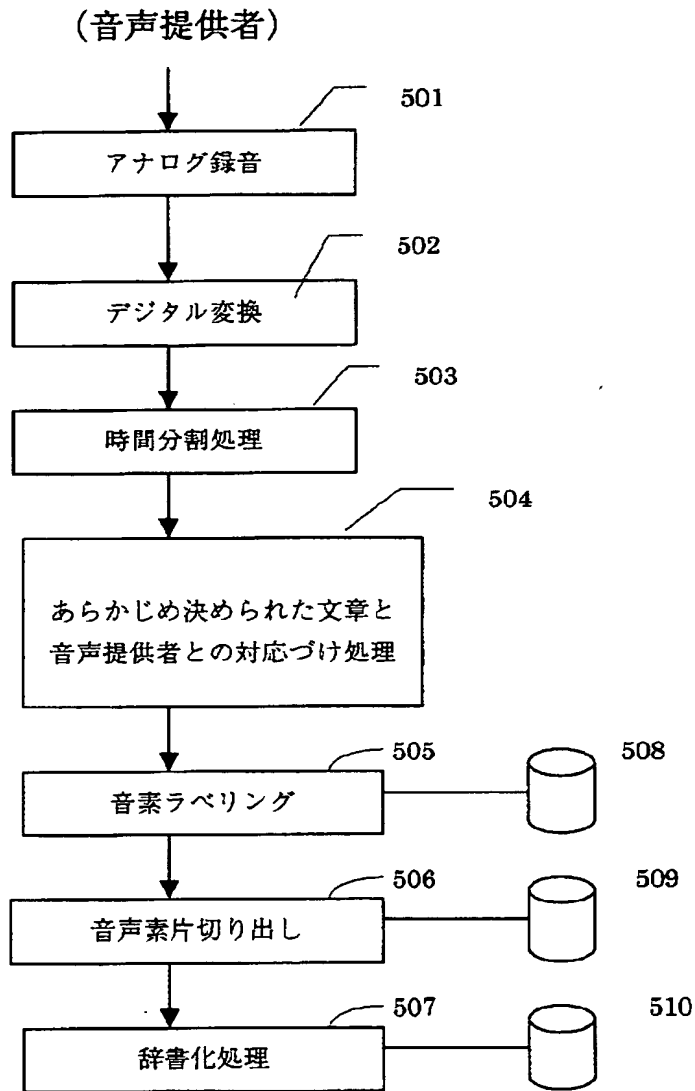
【図11】



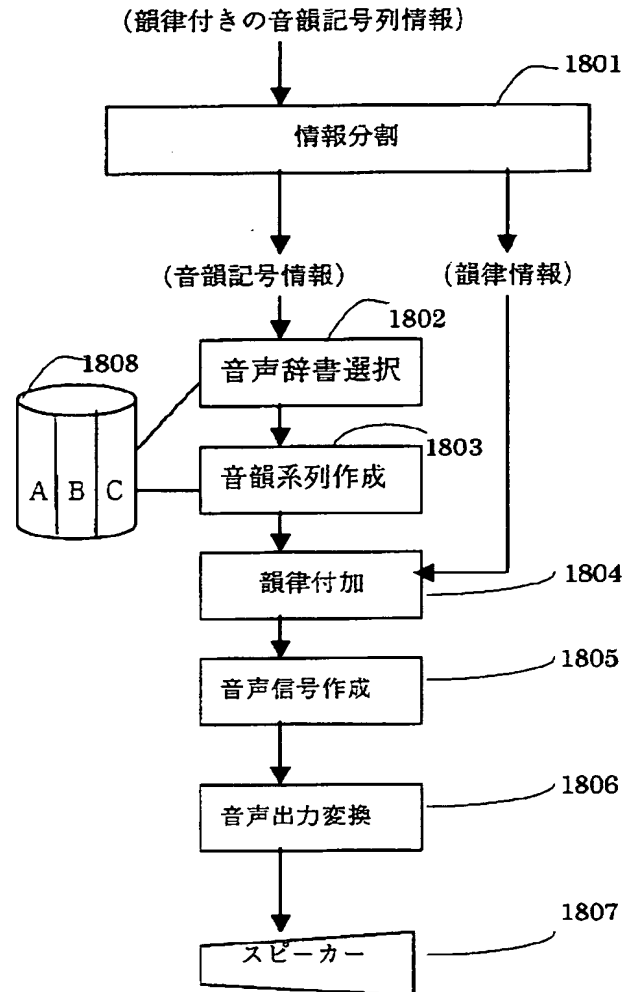
【図16】



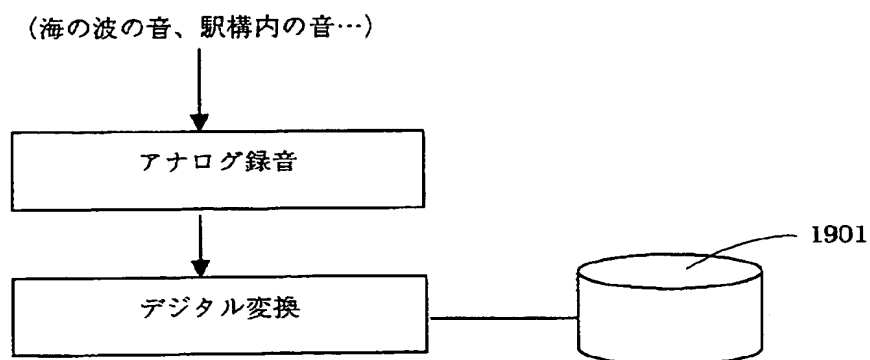
【図 5】



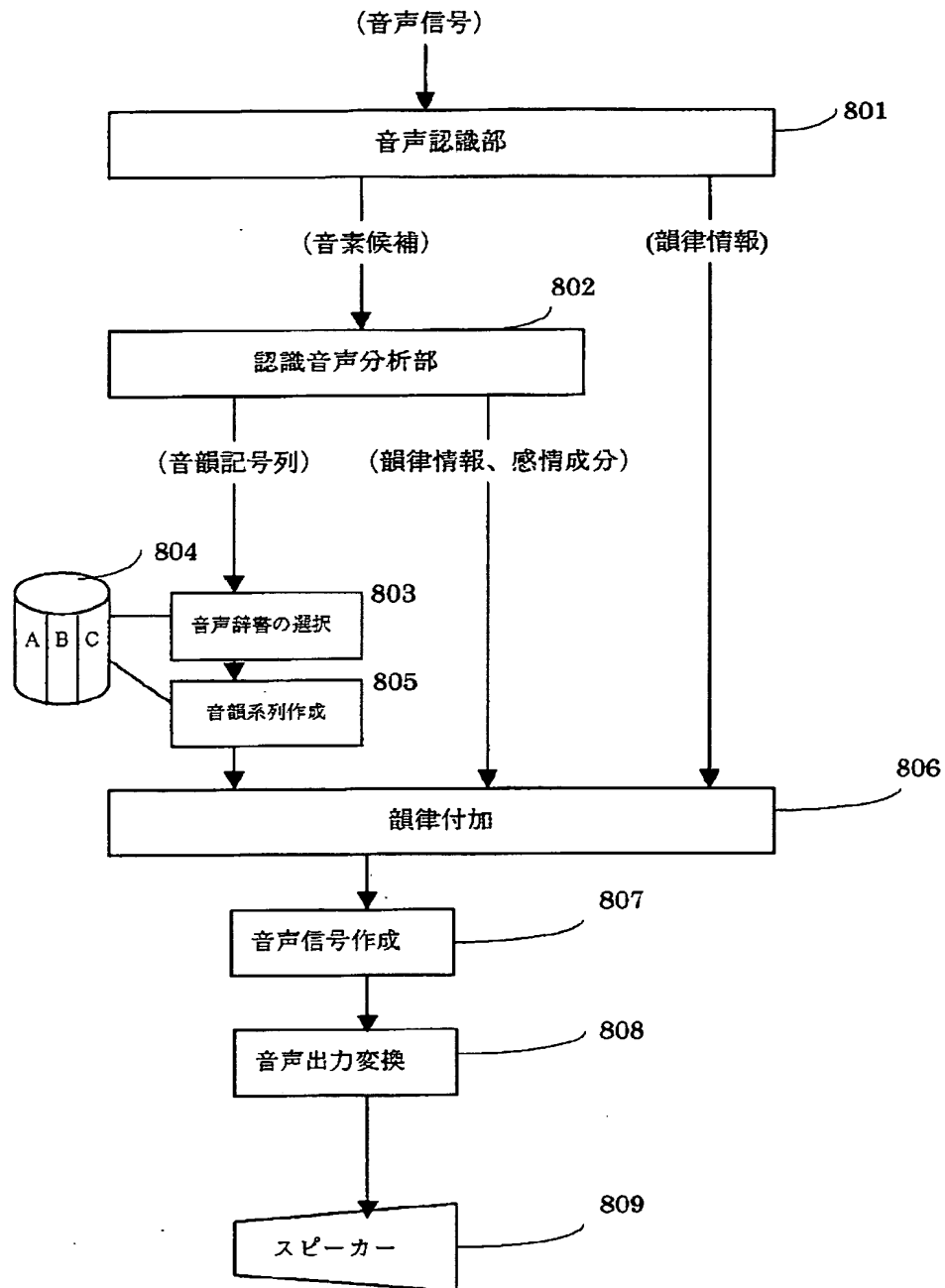
【図 18】



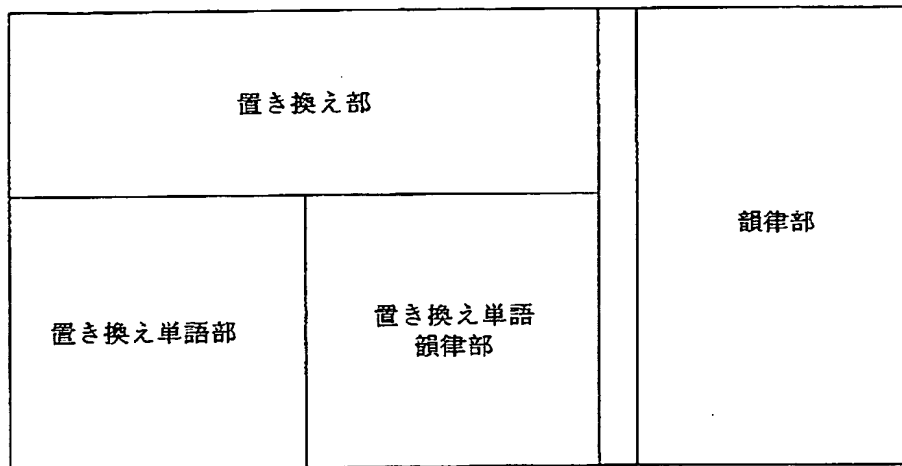
【図 19】



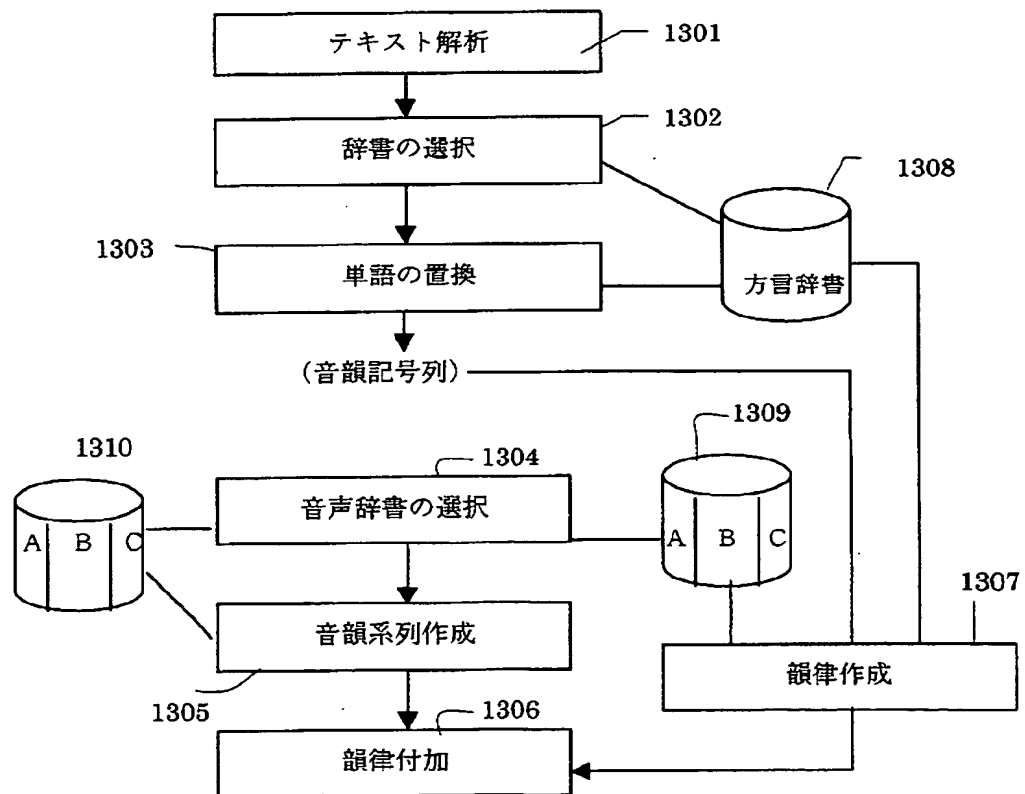
【図8】



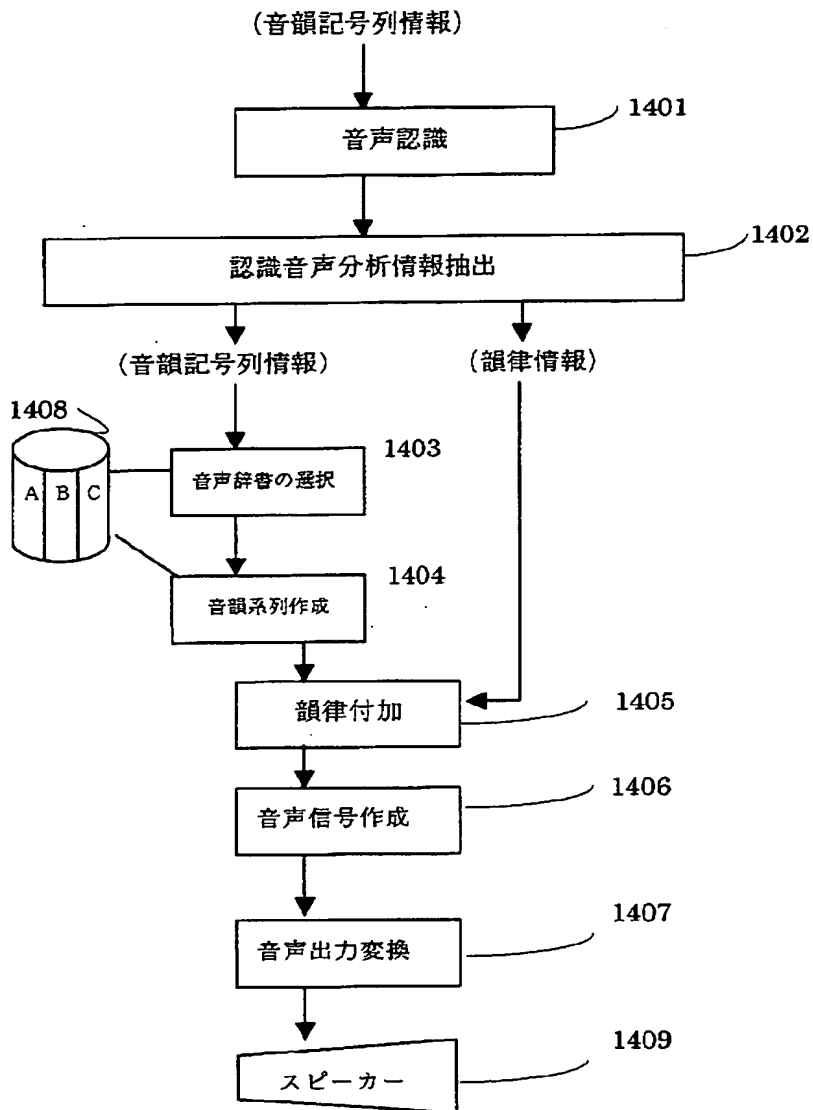
【図 12】



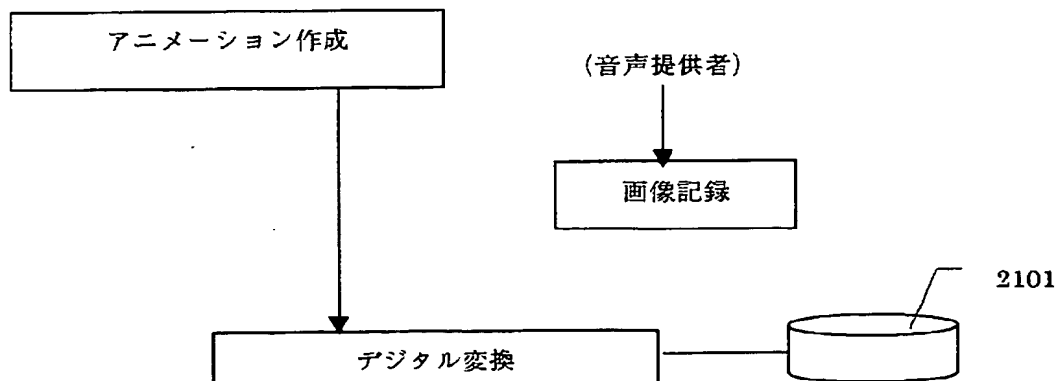
【図 13】



【図14】

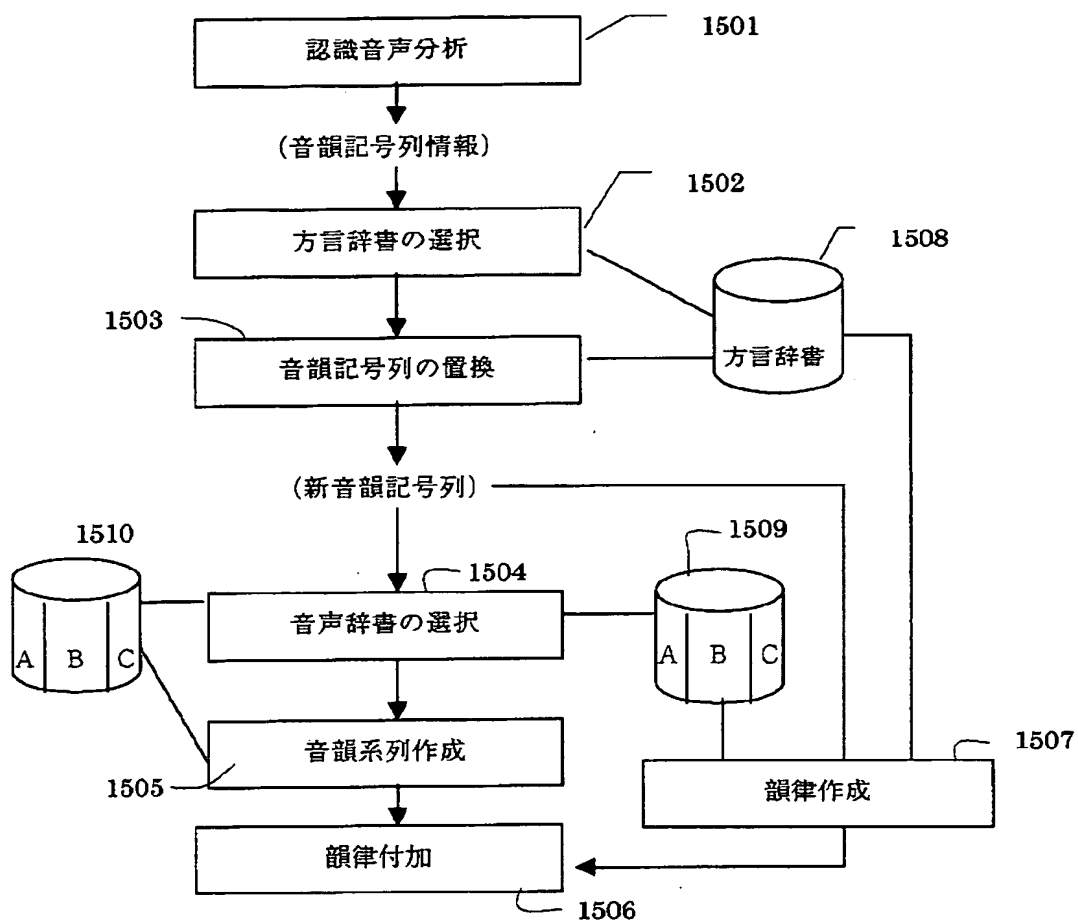


【図21】

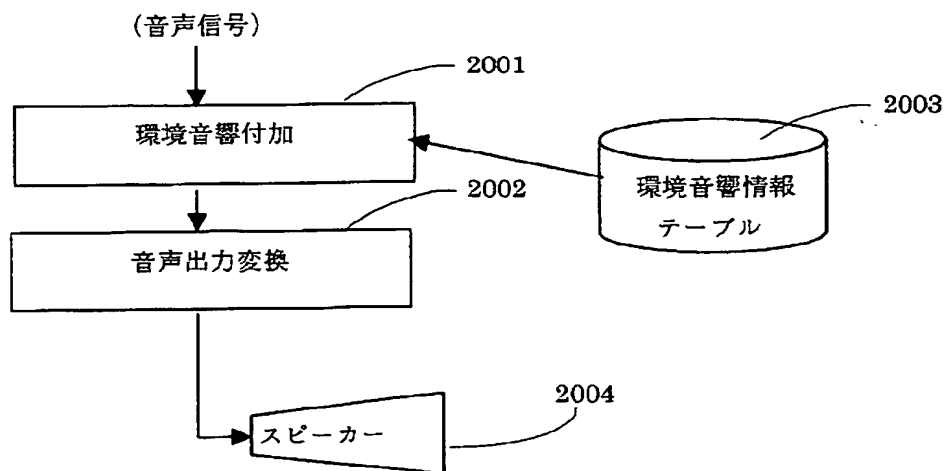




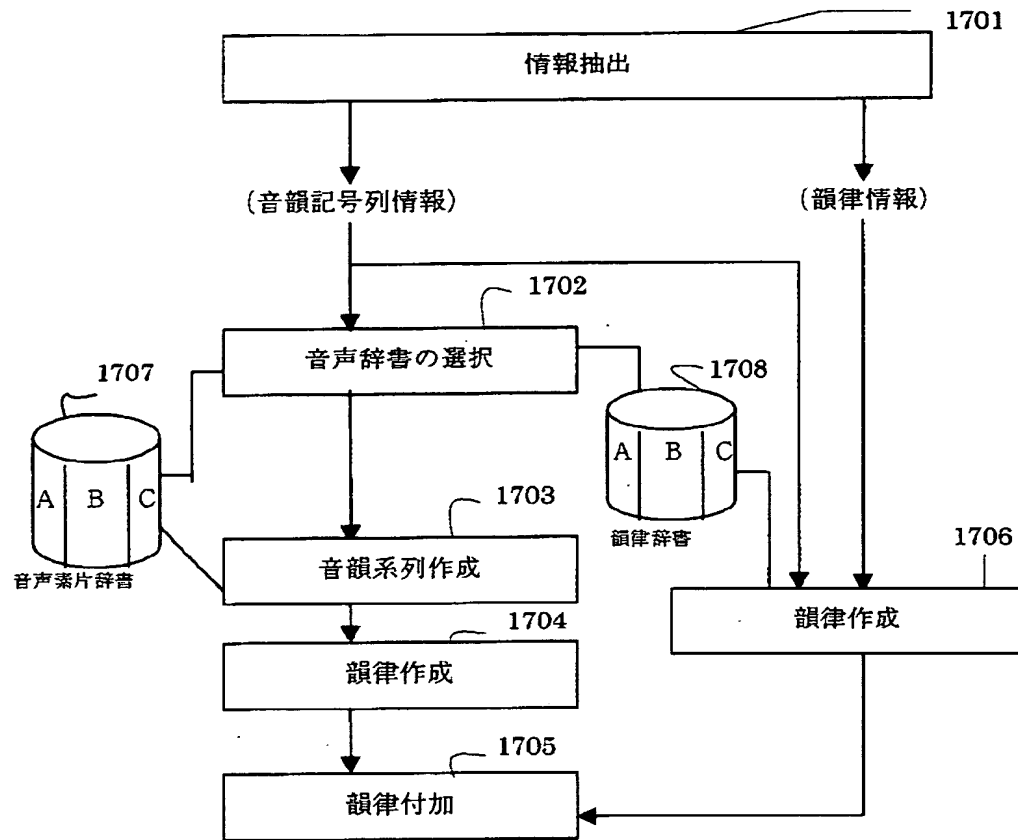
【図 1 5】



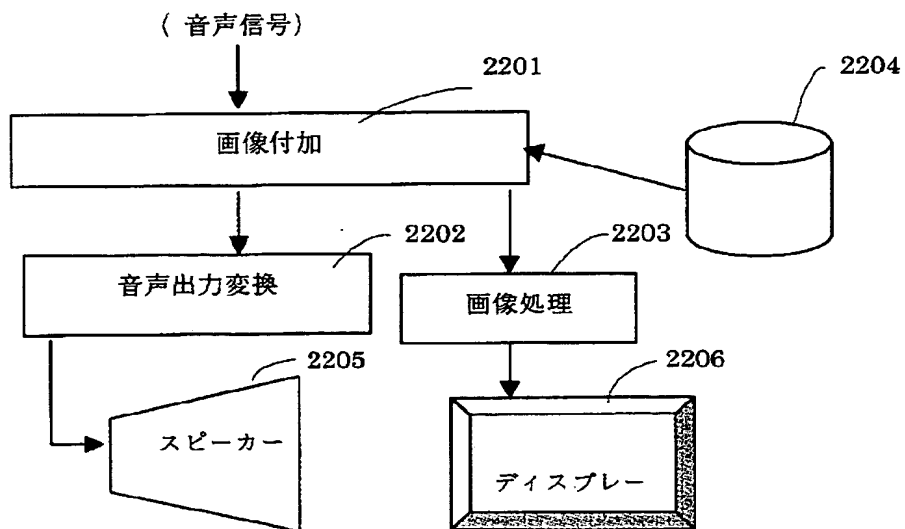
【図 2 0】



【図 17】



【図 22】



フロントページの続き

(51) Int. Cl.<sup>6</sup>

G 0 6 F 3/16

識別記号

3 3 0

F I

G 0 6 F 3/16

3 3 0 K

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**